5   SYSTEM AND METHOD FOR STORING CONNECTIVITY INFORMATION
                           IN A WEB DATABASE

                         ABSTRACT OF THE DISCLOSURE

10

A web crawler system includes a central processing unit for performing computations in
accordance with stored procedures and a network interface for accessing remotely located
computers via a network. A web crawler module downloads pages from remotely located
servers via the network interface. A first link processing module obtains page link
15   information from the downloaded page; the page link information includes for each
downloaded page a row of page identifiers of other pages. A second link processing module
encodes the rows of page identifies in a space efficient manner. It arranges the rows of page
identifiers in a particular order. For each respective row it identifies a prior row, if any, that
best matches the respective row in accordance with predefined row match criteria, determines
20   a set of deletes representing page identifiers in the identified prior row not in the respective
row, and determines a set of adds representing page identifiers in the respective row not in the
identifier prior row. The second link processing module delta encodes the set of deletes and
delta encodes the set of adds for each respective row, and then Huffman codes the delta
encoded set of deletes and delta encoded set of adds for each respective row.

25